

Research article

The PAM domain, a multi-protein complex-associated module with an all-alpha-helix fold

Francesca D Ciccarelli*^{1,2}, Elisa Izaurrealde¹ and Peer Bork^{1,2}Address: ¹European Molecular Biology Laboratory, Meyerhofstr. 1, 69012 Heidelberg, Germany and ²Max-Delbrueck-Centrum, PO Box 740238, D-13092 Berlin, GermanyEmail: Francesca D Ciccarelli* - francesca.ciccarelli@embl.de; Elisa Izaurrealde - elisa.izaurrealde@embl.de; Peer Bork - peer.bork@embl.de

* Corresponding author

Published: 19 December 2003

Received: 21 October 2003

BMC Bioinformatics 2003, 4:64

Accepted: 19 December 2003

This article is available from: <http://www.biomedcentral.com/1471-2105/4/64>

© 2003 Ciccarelli et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Multimeric protein complexes have a role in many cellular pathways and are highly interconnected with various other proteins. The characterization of their domain composition and organization provides useful information on the specific role of each region of their sequence.

Results: We identified a new module, the PAM domain (PCI/PINT associated module), present in single subunits of well characterized multiprotein complexes, like the regulatory lid of the 26S proteasome, the COP-9 signalosome and the Sac3-Thp1 complex. This module is an around 200 residue long domain with a predicted TPR-like all-alpha-helical fold.

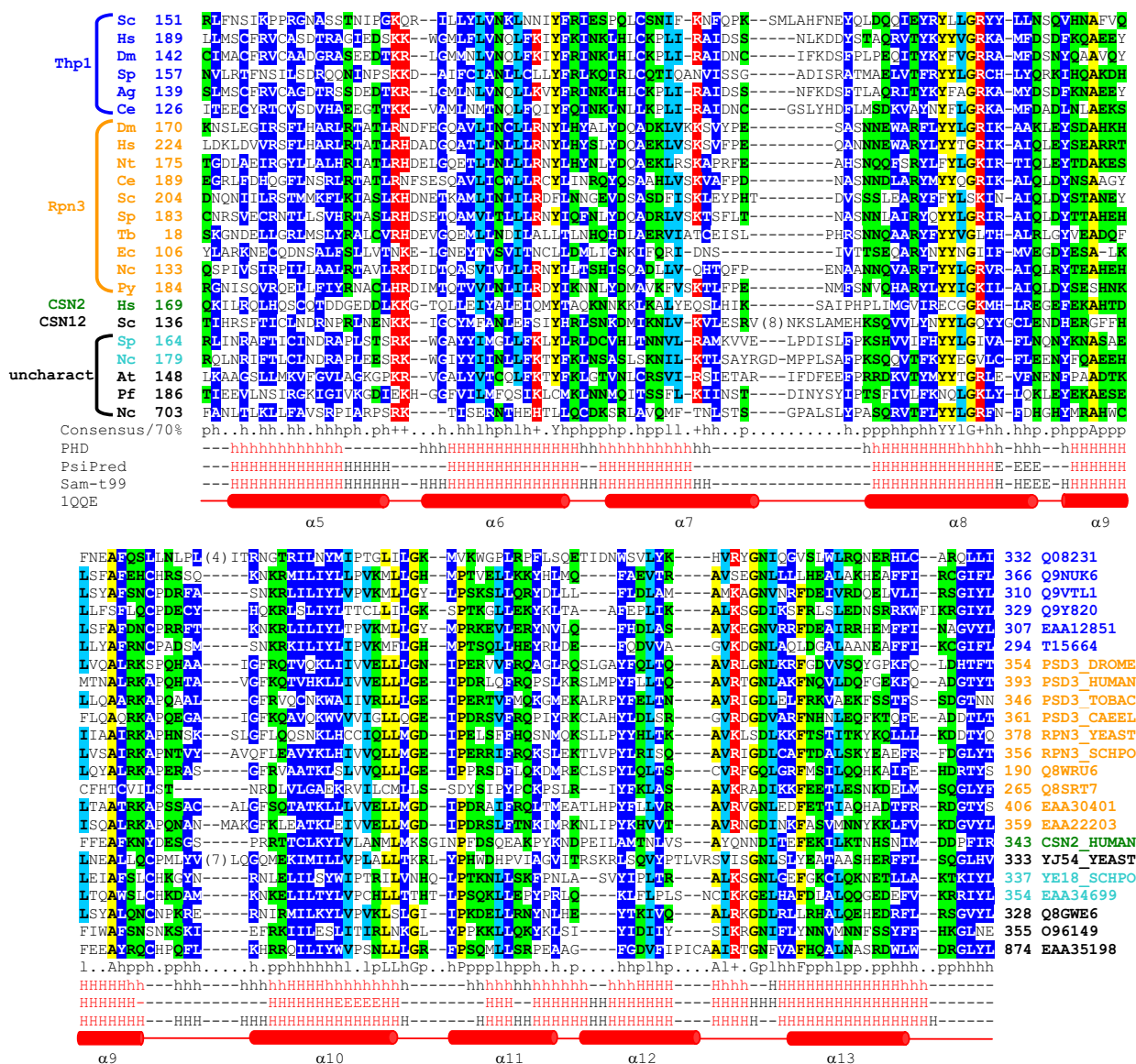
Conclusions: The occurrence of the PAM domain in specific subunits of multimeric protein complexes, together with the role of other all-alpha-helical folds in protein-protein interactions, suggest a function for this domain in mediating transient binding to diverse target proteins.

Background

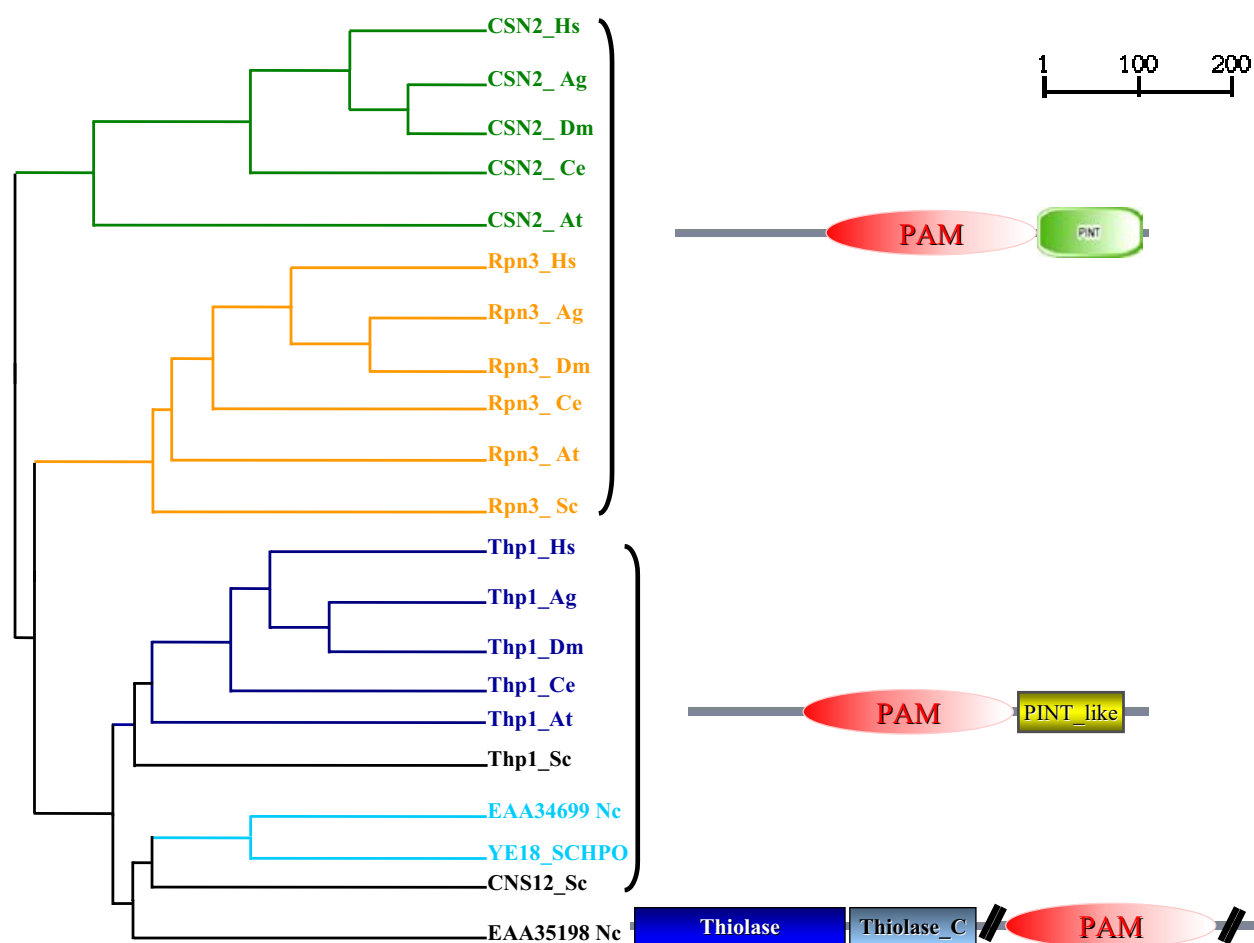
The PCI/PINT (Proteasome, COP9, Initiation factor) and the MPN (Mpr1-Pad1 N-terminal) domains are two modules specifically associated with multiprotein complexes, like the regulatory lid of the 26S proteasome, the COP-9 signalosome (CSN) and the translation initiation factor eIF3 [1,2]. The proteasome regulatory lid and the CSN complexes are composed of the same number of subunits (8) with an identical domain composition, suggesting a common ancestor [3,4]. The gene duplication leading to the two complexes preceded the divergence between unicellular and multicellular organisms and gave rise to two groups of co-orthologous genes [5]. In the case of eIF3 the subunit stoichiometry is not perfectly conserved. The proteasome regulatory lid and eIF3 are well identifiable from yeast to human, whilst the CSN subunits are characterized by a higher degree of divergence in their species distribution. Recently, although not showing a clear one to one

ortholog relationship with the higher eukaryotes' counterparts, the CSN complex has been identified also in *S. cerevisiae* [6,7].

Both the PCI/PINT and the MPN domains undergo rapid changes in their aminoacid composition, likely reflecting their adaptation to specific functions in the different complexes [1,2,8]. In particular, the PCI/PINT domain, a module of around 100 residues, has a predicted α -helical secondary structure but the primary sequence is not well conserved, in particular in the N-terminal part [1,2]. The function of such a domain remains still unclear, although there is evidence of its involvement in directing the incorporation of the subunits in both the proteasome and CSN complexes [9-12].

**Figure 1**

Multiple sequence alignment of representative proteins containing the PAM domain. Of the original set, only sequences with less than 80% identity to each other are shown. The borders of the domain have been assigned taking into account the results of the sequence alignments from PSI-BLAST [17] and the structural alignments from 3D-Jury [23]. Sequences are grouped by phylogenetic relationships, as explained in the legend to Fig. 2. For each sequence, the species, the domain starting and ending residues and the database accession number are reported. The consensus in 70% of the sequences is below the alignment; h, l, p and + indicate hydrophobic, aliphatic, polar, and positive residues, respectively. Hydrophobic residues are highlighted in blue, aliphatic residues in cyan, polar residues in green, positive residues in red and other conserved residues in yellow. The secondary structure predictions using PHD [24], PsiPred [25] and SAM-T99 [26] are reported. For PHD, the upper cases indicate elements predicted with expected average accuracy >82%, and lower cases those predicted with expected average accuracy <82%. The consensus among the three methods is indicated in red. The secondary structure elements of the Sec17 3D structure (1QQE) [27], taken as a representative of the TPR-like structural superfamily, are shown as red cylinders (α-helices: α5–α13). Abbreviations: Ag; *Anopheles gambiae*; At; *Arabidopsis thaliana*; Ce; *Caenorhabditis elegans*; Cs; *Ciona savignyi*; Dm; *Drosophila melanogaster*; Hs; *Homo sapiens*; Nc; *Neurospora crassa*; Nt; *Nicotiana tabacum*; Pf; *Plasmodium falciparum*; Py; *Plasmodium yoelii*; Sc; *saccharomyces cerevisiae*; Sp; *Schizosaccharomyces pombe*; Tb; *Trypanosoma brucei*, Ec; *Encephalitozoon cuniculi*, H, helix.

**Figure 2**

Phylogenetic tree and domain architecture of the protein families bearing the PAM domain. The tree was built using the corresponding PAM domain region of the CSN2, Rpn3, Thp1 proteins in representative species. Where uncharacterized, the proteins are indicated with their database accession number. The nodes with a bootstrap support exceeding 90% are indicated with coloured lines, those with a bootstrap value below 90% are reported in black. The domain architecture was derived from the SMART database [28]. A divergent version of the PCI/PINT domain is indicated in yellow (Table 1, see 1). It should be noted that the *N. crassa* EAA35198 protein is a hypothetical protein derived from gene prediction. As it is the only protein with a PAM domain showing the thiolase domains, it could be that the gene prediction is wrong and the actual protein is bearing only the PAM domain. Abbreviations: PINT: Proteasome, Int-6, Nip-1 and TRIP-15; PAM, PCI/PINT associated module; Thiolase: Thiolase, N-terminal domain; Thiolase_C: Thiolase, C-terminal domain.

Results and Discussion

In the course of the characterization of the different subgroups inside the PCI/PINT domain family, we identified a new region of sequence similarity associated with PCI/PINT domains, which we called PAM (for PCI/PINT associated module, Fig. 1). The starting point of our analysis was *S. cerevisiae* Thp1 (Q08231), a protein involved in transcription regulation and messenger RNA export [13-16]. In the C-terminal part of this protein (roughly from

residue 300 to residue 430), a divergent PCI/PINT domain can be detected (Table 1, see 1), whilst in the N-terminal part no known modules are annotated. When using the entire protein to search the non-redundant protein database (nrdb), the region preceding the PCI/PINT domain showed significant similarity to uncharacterized proteins. We then restricted our search only to the N-terminal part of the sequence (residues 1 to 300). At the first iteration of PSI-BLAST [17], the putative orthologs of

Thp1 in *H. sapiens*, *D. melanogaster*, *A. gambiae* and *S. pombe* were retrieved (Q9NUK6, Q9VTL1, EAA12851 and Q9Y820, respectively, with E-values from 10^{-05} to 10^{-03}). After detecting uncharacterized proteins from *A. thaliana* (Q8GWE6, $E = 10^{-15}$), *N. crassa* (EAA34699, $E = 10^{-15}$ and EAA35198, $E = 10^{-08}$), *S. pombe* (YE18_SCHPO, $E = 10^{-11}$) and the CSN12 subunit of the yeast signalosome (YJ54_YEAST, $E = 10^{-09}$), the search converged. A HMM profile derived from the non-redundant multiple alignment of the new region was used to refine the search in the nrdb with HMMer [18]. This analysis added the orthologs of the Rnp3 proteasome subunit and some uncharacterized proteins to the initial sequence-set. The final non-redundant multiple alignment of representative members of all the families is shown in Fig. 1. Interestingly, none of the eIF3 subunits was detected.

The PAM domain is around 200 residue long, with a recursive occurrence of hydrophobic patches followed by conserved positive residues. This periodicity in the amino acid composition is typical for a structure rich in α -helices. Indeed, several secondary structure prediction methods indicate helical elements all along the domain (Fig. 1). These results have been corroborated by the fold predictions obtained using as queries different sequences of the multiple alignment. In all cases the SCOP [19] TPR-like superfamily has been indicated as the most likely fold for the new domain (Table 2, see 1). This superfamily is composed of all-helical structures, like TPRs (tetrapeptide repeats) and HLH (helix-loop-helix) domains. The structural similarity to TPRs is also confirmed by the PFAM database [20], which, in the case of the Rnp3 subunits, predicts a single TPR covering a small region inside their PAM domain. In this particular case, the PFAM prediction has the value of a structural indication more than the detection of a real TPR. The PAM domain is indeed much larger than a TPR and it is not possible to identify any clear repeat inside this region using different resources, as REP [21] and ARIADNE [22]. Therefore the PAM domain is a distinct α -helical module specifically occurring in a subset of PCI/PINT domain containing proteins. Notably, while the PCI/PINT domains show high divergence, the PAM modules are highly conserved among the sequences (Fig. 1) supporting the existence of two independent domains.

Conclusions

Different structural domains composed of α -helical elements, including TPR, HEAT, armadillo and clathrin heavy chain repeats, are characterized by a superhelical arrangement of repeats, which eventually results in a binding surface often mediating interactions with other proteins. The structural indication of a TPR-like fold and the presence of the PAM domain in some of the constituents of characterized multiprotein complexes suggest an

involvement of the domain in mediating protein-protein interaction. Interestingly, unlike the PINT/PCI and the MPN domains occurring in many subunits, the PAM domain is detectable in only one of the subunits of both the proteasome lid and the CSN complexes (Rpn3 and CSN2, respectively, Fig. 2), but in no subunit of the eIF3 complex. This indicates a more specific role for the PAM domain in mediating transient interactions to proteins others than the complex components. In particular in the case of CSN2, several of such transient interactions have been reported [5,12].

The *S. cerevisiae* Thp1 protein is a component of the Sac3-Thp1 complex, which was primarily found to have a role in transcription [13,15]. Recent data showed an involvement of the Sac3-Thp1 complex in mRNA export from the nucleus to the cytoplasm and its interaction with other multiprotein complexes of the same pathway [14,16].

In summary, the PAM domain is an α -helical module present in single subunits of multimeric complexes. As other domains in the same complexes (e.g. the PCI/PINT domain) have been proposed to mediate the internal interactions between the subunits, the PAM domain might play a role in the transient binding to the different external targets.

Authors' contributions

EI furnished the initial input to the research. FDC carried out the sequence analysis and the domain characterization. FDC, EI and PB authored the manuscript.

Additional material

Additional File 1

Additional data are available at http://www.bork.embl.de/~ciccarel/PAM_add_data.pdf. Table 1: Results of the HMM search using the PINT HMM profile against a set of PAM-containing sequences. Table 2: Results of the structure predictions obtained using the metasever 3D-Jury. Click here for file
[<http://www.biomedcentral.com/content/supplementary/1471-2105-4-64-S1.pdf>]

Aknowledgments

The authors are grateful to the members of the Bork group for the useful comments on the manuscript.

References

1. Aravind L, Ponting CP: **Homologues of 26S proteasome subunits are regulators of transcription and translation.** *Protein Sci* 1998, **7**:1250-1254.
2. Hofmann K, Bucher P: **The PCI domain: a common theme in three multiprotein complexes.** *Trends in Biochemical Sciences* 1998, **23**:204-205.
3. Glickman MH, Rubin DM, Coux O, Wefes I, Pfeifer G, Cjeka Z, Baumeister W, Fried VA, Finley D: **A subcomplex of the proteasome**

- regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* 1998, **94**:615-623.
4. Kim T-H, Hofmann K, von Arnim AG, Chamovitz DA: **PCI complexes: pretty complex interactions in diverse signaling pathways.** *Trends in Plant Science* 2001, **6**:379-386.
 5. Serino G, Su H, Peng Z, Tsuge T, Wei N, Gu H, Deng XW: **Characterization of the Last Subunit of the Arabidopsis COP9 Signalosome: Implications for the Overall Structure and Origin of the Complex.** *Plant Cell* 2003, **15**:719-731.
 6. Maytal-Kivity V, Piran R, Pick E, Hofmann K, Glickman MH: **COP9 signalosome components play a role in the mating pheromone response of *S. cerevisiae*.** *EMBO Rep* 2002, **3**:1215-1221.
 7. Maytal-Kivity V, Pick E, Piran R, Hofmann K, Glickman MH: **The COP9 signalosome-like complex in *S. cerevisiae* and links to other PCI complexes.** *The International Journal of Biochemistry & Cell Biology* 2003, **35**:706-715.
 8. Maytal-Kivity V, Reis N, Hofmann K, Glickman M: **MPN+, a putative catalytic motif found in a subset of MPN domain proteins from eukaryotes and prokaryotes, is critical for Rpn11 function.** *BMC Biochemistry* 2002, **3**:28.
 9. Kapelari B, Bech-Otschir D, Hegerl R, Schade R, Dumdey R, Dubiel W: **Electron Microscopy and Subunit-Subunit Interaction Studies Reveal a First Architecture of COP9 Signalosome.** *Journal of Molecular Biology* 2000, **300**:1169-1178.
 10. Tsuge T, Matsui M, Wei N: **The Subunit I of the COP9 Signalosome Suppresses Gene Expression Through its N-terminal Domain and Incorporates into the Complex Through the PCI Domain.** *Journal of Molecular Biology* 2001, **305**:1-9.
 11. Fu H, Reis N, Lee Y, Glickman MH, Vierstra RD: **Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome.** *EMBO J* 2001, **20**:7096-7107.
 12. Lier S, Paululat A: **The proteasome regulatory particle subunit Rpn6 is required for *Drosophila* development and interacts physically with signalosome subunit Alien/CSN2.** *Gene* 2002, **298**:109-119.
 13. Gallardo M, Aguilera A: **A New Hyperrecombination Mutation Identifies a Novel Yeast Gene, THP1, Connecting Transcription Elongation With Mitotic Recombination.** *Genetics* 2001, **157**:79-89.
 14. Gallardo M, Luna R, Erdjument-Bromage H, Tempst P, Aguilera A: **Nab2p and the Thp1p-Sac3p Complex Functionally Interact at the Interface between Transcription and mRNA Metabolism.** *J Biol Chem* 2003, **278**:24225-24232.
 15. Gonzalez-Barrera S, Garcia-Rubio M, Aguilera A: **Transcription and Double-Strand Breaks Induce Similar Mitotic Recombination Events in *Saccharomyces cerevisiae*.** *Genetics* 2002, **162**:603-614.
 16. Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E: **The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores.** *EMBO J* 2002, **21**:5843-5852.
 17. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
 18. Eddy SR: **Profile hidden Markov models.** *Bioinformatics* 1998, **14**:755-763.
 19. Lo Conte L, Brenner SE, Hubbard TJ, Chothia C, Murzin AG: **SCOP database in 2002: refinements accommodate structural genomics.** *Nucleic Acids Res* 2002, **30**:264-267.
 20. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer ELL: **The Pfam Protein Families Database.** *Nucl Acids Res* 2002, **30**(1):276-280.
 21. Andrade MA, Ponting CP, Gibson TJ, Bork P: **Homology-based method for identification of protein repeats using statistical significance estimates.** *Journal of Molecular Biology* 2000, **298**:521-537.
 22. Mott R: **Accurate formula for P-values of gapped local sequence and profile alignments.** *Journal of Molecular Biology* 2000, **300**:649-659.
 23. Ginalski K, Elofsson A, Fischer D, Rychlewski L: **3D-Jury: a simple approach to improve protein structure predictions.** *Bioinformatics* 2003, **19**:1015-1018.
 24. Rost B: **PHD: predicting one-dimensional protein structure by profile-based neural networks.** *Methods Enzymol* 1996, **266**:525-539.
 25. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16**:404-405.
 26. Karplus K, Barrett C, Hughey R: **Hidden Markov models for detecting remote protein homologies.** *Bioinformatics* 1998, **14**:846-856.
 27. Rice LM, Brunger AT: **Crystal structure of the vesicular transport protein Sec17: implications for SNAP function in SNARE complex disassembly.** *Mol Cell* 1999, **4**:85-95.
 28. Letunic I, Goodstadt L, Dickens NJ, Doerks T, Schultz J, Mott R, Ciccarelli F, Copley RR, Ponting CP, Bork P: **Recent improvements to the SMART domain-based sequence annotation resource.** *Nucleic Acids Res* 2002, **30**:242-244.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

